ANALES DE LA REAL ACADEMIA NACIONAL DE FARMACIA
ISSN (Online) 1697-4298
Opinion Pieces
analesranf.com

Reflections about the scientific process

Title in Spanish: *Reflexiones sobre el proceso científico*

Carmen Avendaño[1,*]

[1]Académica de Número de la Real Academia Nacional de Farmacia, Madrid.

[*]**Corresponding Author:** avendano@farm.ucm.es

An Real Acad Farm Vol. 83, Nº 1 (2017), pp. 6-9
Received: February 27, 2017 **Accepted:** February 27, 2017
Language of Manuscript: Spanish

1. METACIENCIA

Recientemente, la política nos ha familiarizado con la *post-truth* ("posverdad" o "metaverdad"), un nuevo concepto que suele referirse a campañas electorales que recurren a las emociones prescindiendo de la verdad de los hechos. El prefijo "meta" tiene en este caso uno de los significados que le da el Diccionario de la Real Academia Española: "después de" o "más allá", pero este prefijo también se aplica a términos como "metalenguaje", significando que el concepto que designa el sustantivo recae sobre sí mismo (un lenguaje que reflexiona sobre el lenguaje mismo). Análogamente, la metaciencia puede definirse como el estudio científico de la ciencia.

Esta materia está floreciendo en los últimos años, y sus trabajos han evidenciado que la imagen idealizada del modelo hipotético-deductivo del método científico, cuyo sello de identidad es su capacidad para descubrir a partir de unos datos experimentales patrones nuevos e inesperados, está amenazada por varios factores, lo que afecta a la eficacia del conocimiento que la ciencia proporciona. Uno de estos factores es la falta de reproducibilidad de los resultados científicos que se publican, lo que ocurre con frecuencia en las ciencias de la vida como veremos en algunos ejemplos que hemos seleccionado.

2. LA REPRODUCIBILIDAD DE LOS RESULTADOS EXPERIMENTALES

En el año 2009, cuatro grupos de analistas liderados por John Ioannidis (un médico e investigador de la Universidad de Stanford que analiza la calidad del trabajo de otros científicos), intentaron reproducir 18 estudios sobre la expresión genética utilizando chips de ADN que se habían publicado en *Nature Genetics* los años 2005 y 2006. Utilizando los datos y los métodos analíticos detallados en dichas publicaciones concluyeron que solo habían sido capaces de reproducir 2 de ellos, otros 6 pudieron reproducirse con algunas discrepancias y 10 no pudieron ser reproducidos. Los autores achacaron este fracaso a la falta de algunos datos y a que no se especificaba cómo se habían procesado y analizado los datos publicados (1). Tres años más tarde, varios investigadores de la compañía biotecnológica Amgen en Thousand Oaks (California), solo pudieron reproducir 6 de 53 estudios de oncología y hematología que habían sido publicados, concluyendo que el poco éxito logrado en los ensayos clínicos con posibles fármacos antitumorales se debía a deficiencias en el planteamiento de los ensayos preclínicos (uso de líneas celulares y modelos animales inadecuados), además de a las dificultades de esta enfermedad asociadas a la heterogeneidad de los tumores y de los pacientes. Los autores proponían que, para beneficiar a los pacientes de cáncer, deberían cambiarse los métodos, las publicaciones y los incentivos (2). En el año 2016, el 90% de los 1.500 investigadores que respondieron a una encuesta de la revista *Nature*, estuvieron de acuerdo en que existe una crisis de reproducibilidad (3).

Es evidente que la baja reproducibilidad de las investigaciones básicas y preclínicas en el campo de las ciencias de la vida, contribuye a un retraso y a un mayor coste en el desarrollo de fármacos. Los recursos que se emplean en el diseño de los experimentos para minimizar el sesgo de los resultados en los estudios clínicos (como la aleatorización y la evaluación por investigadores "ciegos" que desconocen las condiciones experimentales), no suelen darse en los estudios básicos o preclínicos, lo que conduce a que el evaluador tienda a encontrar los resultados que está buscando e ignore los que le resultan incoherentes. Según algunos analistas, más de un 50% de las investigaciones preclínicas publicadas en EEUU son irreproducibles, lo que significa un coste para este país de 28.000.000.000 dólares/año (4).

Este problema tiene distintas causas, entre las que se encuentran el que se propongan las hipótesis cuando ya se conocen los resultados (5), que se practique el *P-hacking* y se aumente la flexibilidad en el análisis de los datos, y que no se compartan datos que pueden ser relevantes.

Según los analistas, aunque la ciencia no es inmune a las conductas fraudulentas, la falta de reproducibilidad de los resultados que se publican no va ligada necesariamente a que los resultados hayan sido fabricados, ya que puede deberse a la variabilidad biológica, a problemas en el

diseño de los experimentos o a fallos en el análisis de los resultados obtenidos. El comportamiento de los sistemas biológicos es muy complejo y se ve afectado por factores ambientales, especialmente en el campo de las neurociencias (6). Finalmente hay que hablar de la estadística, que es clave en el análisis de los resultados de una investigación pero que se utiliza algunas veces para "descubrir" falsos positivos, correspondientes a hallazgos que aparecen como significativos cuando en realidad no los son (7).

3. LA SIGNIFICACIÓN ESTADÍSTICA

En estadística, una hipótesis nula (o hipótesis de partida) es una afirmación que no se rechaza a menos que los datos de la muestra en estudio parezcan evidenciar que es falsa. El valor "P" es una medida de significación estadística que se define como la probabilidad de obtener un resultado al menos tan extremo como el que realmente se ha obtenido suponiendo que la hipótesis nula es cierta. Cuanto menor sea este valor, mayor será la significación estadística de los resultados experimentales. Si el valor P asociado al resultado observado es menor que el nivel de significación establecido convencionalmente (frecuentemente $P < 0,05$ en investigaciones sociológicas y $P < 0,01$ en investigaciones clínicas), lo más verosímil es que la hipótesis de partida sea falsa y que los resultados encontrados experimentalmente sean reales y no debidos al azar, soportando por tanto una hipótesis correcta.

Es posible que si se encuentra una observación atípica se rechace la hipótesis nula aunque ésta sea cierta, pero este error estadístico puede subsanarse rebajando el valor P o aumentando el tamaño de la muestra para reducir la posibilidad de que el dato obtenido sea casualmente raro. Sin embargo, también es posible que se elimine alguna condición experimental a fin de que el valor de P tenga significación estadística, o no se incluyan en la publicación los datos sin procesar. Estas maniobras, realizadas en el análisis de los datos quizás de forma inconsciente, se conocen como "*P-hacking*", y podrían evitarse si se proponen las comparaciones antes de que se realicen las experiencias y se detallen en las publicaciones los métodos y el análisis de datos, incluyendo los que difieren de lo que se planificó.

Las trabajos relacionados con la credibilidad de los resultados científicos aportan conclusiones demoledoras (8) que, en los estudios biomédicos, afectan tanto a ensayos clínicos y a estudios epidemiológicos (9) como a las investigaciones realizadas a nivel molecular (10). Pero el problema afecta también a otras disciplinas científicas.

La metaciencia demuestra que, además de el P-hacking, la publicación de resultados incorrectos puede producirse por otros factores, como son el estudio de muestras de pequeño tamaño, la medición de efectos pequeños, o los conflictos de interés. Estos últimos no tienen que estar necesariamente ligados al problema de la financiación o al rendimiento económico de las investigaciones, sino que pueden deberse al convencimiento de los investigadores de que una teoría determinada es cierta o a la necesidad de producir resultados para mantener una situación profesional o promocionarse, lo que se da frecuentemente en los centros de investigación financiados públicamente, ya que el entorno académico es hoy más competitivo que nunca (11).

El sesgo de publicación induce a los investigadores a enviar para su publicación aquellas investigaciones que arrojen resultados "positivos" (hallazgos significativos) en lugar de "negativos" ("que apoyen la hipótesis nula"). Sólo si un asunto es de actualidad y otro equipo acaba de publicar un resultado positivo, los resultados negativos acerca de la misma cuestión pueden resultar muy "atractivos". Este sesgo también podría aplicarse a los editores, que acepten más fácilmente publicaciones con resultados positivos, por considerarlos más interesantes que los negativos.

4. INTENTOS PARA ATAJAR EL PROBLEMA

En los últimos años, la falta de reproducibilidad en los resultados científicos se trata de atajar a través de distintas iniciativas. El *National Institute of Health* (NIH) de los EEUU y los responsables de las revistas *Nature* y *Science* reunieron en Junio de 214 a varios editores. Éstos representaban a unas 30 revistas de ciencia básica y preclínica en donde se publican la mayor parte de las investigaciones financiadas por el NIH. Tras analizar cómo podía aumentarse el rigor científico y lograr que la ciencia sea reproducible, sólida y transparente, se consensuaron varias recomendaciones que incluían realizar un análisis estadístico riguroso (lo que requiere una gran formación estadística en los investigadores), que la información que se suministre sea transparente, que se compartan el material y los datos, y que se tengan en cuenta los datos en contra (12). Ese mismo año, antes del encuentro anual de la *Society for Neuroscience* de EEUU, el entonces Director del *National Institute of Mental Health* (NIMH) Thomas Insel constató que los estudios sobre la estructura y funciones del cerebro habían experimentado un progreso extraordinario, pero advirtió que no podía olvidarse que hasta un 80% de los datos que dichos estudios habían proporcionado no habían podido reproducirse. En palabras de T. Insel *"Great science requires great attention to the details of experimental design and data analysis"*.

Recientemente, el 10 de enero de 2017, un grupo de investigadores liderado por John Ioannidis ha publicado un manifiesto para que la ciencia recupere la credibilidad y fiabilidad y se aceleren los descubrimientos (13). Según este manifiesto, el 85% de los esfuerzos dedicados a investigaciones biomédicas se abandonan en etapas muy tempranas o no llegan a aplicarse en clínica, y señala que tanto los investigadores jóvenes como los *senior* requieren una educación metodológica continuada, especialmente en el uso del análisis estadístico. Sin embargo, a muchos científicos les intimida esta herramienta, y proponen que se desmitifique y se señale un marco alternativo para utilizar los modelos matemáticos que se requieran (14).

5. ALMACENAMIENTO Y ANÁLISIS DE DATOS MASIVOS (*BIG DATA* O MACRODATOS)

El término *big data* (macrodatos, datos masivos o datos a gran escala), hace referencia al almacenamiento de grandes cantidades de datos y a los procedimientos usados para encontrar patrones repetitivos dentro de ellos. Estas técnicas, que surgieron de la necesidad de confeccionar informes estadísticos y modelos predictivos, representan una revolución en las tecnologías de la información y la comunicación, y están cambiando la forma de hacer negocios, la política, la educación, la sanidad, y la innovación (15), constituyendo una oportunidad para la investigación. Los científicos necesitan hoy día complementar la ciencia con las búsquedas inteligentes de datos, cuyo volumen crece constantemente, y adaptar sus prácticas a esta nueva herramienta. Su utilización es también un reto, ya que estos datos han de procesarse inteligentemente (16). No puede perderse de vista la enorme diferencia que existe entre los *big data* científicos, que se extraen de muchos trabajos experimentales, y los *big data* extraidos de conductas humanas, ya que estos últimos son muchas veces contradictorios.

Entre las muchísimas herramientas para analizar los *big data*, se encuentra su agrupación (*clustering*), en la que grandes grupos de datos dan lugar a grupos más pequeños en función de su semejanza, desconocida antes de realizar este tipo de análisis. Esta técnica, referida a individuos, se está aplicando en el campo de la salud. Por ejemplo, utilizando los datos de búsquedas que contenían los términos *Influenza-Like Illness Symptoms*, agregados según ubicación y fecha, *Google Flu Trends*, predijo hacia mediados de 2009 una pandemia de gripe A, con dos semanas de antelación a los sistemas de detección tradicionales. El análisis de datos masivos ha comenzado ya a aplicarse en el tratamiento del cáncer y en el diagnóstico de enfermedades, campos en los que la inteligencia artificial es capaz de realizar análisis con una gran precisión (17).

6. CONCLUSIÓN

La metaciencia está demostrando que es posible hacer una investigación más exigente y reproducible (18), y descubrir que lo que considerábamos como verdadero puede que no lo sea. En palabras de Regina Nuzzo: *Humans are remarkably good at self-deception, but growing concern about reproducibility is driving many researchers to seek ways to fight their own worst instincts* (20). Fomentar la autocorrección del propio proceso científico puede ser de gran utilidad en medicina, tanto en lo que se refiere a métodos diagnósticos (19) como a tratamientos.

7. BIBLIOGRAFÍA

1. Ioannidis JPA, Allison DB, Ball CA, *et al*. Repeatability of published microarray gene expression analyses. Nature Genet 2009; 41: 149-55.
2. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature 2012; 483: 531-3.
3. Baker M. 1,500 scientists lift the lid on reproducibility. Nature 2016; 533: 452-4.
4. Freedman L, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. PLoS Biology 2015; 13: e1002165.
5. Kerr NL. HARKing: hypothesizing after the results are known. Pers Soc Psychol Rev 1998; 2: 196-217.
6. Crabbe JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. Science 1999; 284: 1670-2.
7. Simmons JP, Nelson LD, Simonshohn U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci 2011; 22: 1359-66.
8. a) Ioannidis JPA. Why most published research findings are false. PLoS Med 2005; 2, e124. b) Fanelli D. "Positive" results increase down the Hierarchy of the Sciences. PloS ONE 2010; 5: e10068. c) Button KS, Ioannidis JPA, Mokrysz C, *et al*. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013; 14: 365-76.
9. a) Ioannidis JP, Haidich AB, Lau J. Any casualties in the clash of randomised and observational evidence? BMJ 2001; 322: 879-80. b) Lawlor DA, Smith GD, Kundu D, *et al*. Those confounded vitamins: What can we learn from the differences between observational versus randomised trial evidence? The Lancet 2004; 363: 1724-7. c) Vandenbroucke JP. When are observational studies as credible as randomised trials? The Lancet 2004; 363: 1728-31.
10. a) Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: A multiple random validation strategy. The Lancet 2005; 365: 488-92. b) Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001; 29: 306-9.
11. Papanikolaou GN, Baltogianni MS, Contopoulos-Ioannidis DG, *et al.* Reporting of conflicts of interest in guidelines of preventive and therapeutic interventions. BMC Med Res Methodol 2001; DOI: 10.1186/1471-2288-1-3.
12. a) Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature 2014; 505: 612-3. b) Motulsky HJ. Common misconceptions about data analysis and statistics. J Pharmacol Exp Ther 2014; 351: 200-5.
13. Munafò MR, Nosek BA, Bishop DVM, Button KS, *et al*. A manifesto for reproducible science. Nature Human Behaviour 2017, doi:10.1038/s41562-016-0021. 14. Miller MJ, van den Heuvel ER, Roesti D. The role of statistical analysis in validating rapid microbiological methods. European Pharmaceutical Review RMMs Suplement, 2016; issue 6.
15. Mayer-Schönberger V, Cukier K. Big Data: A Revolution That Will Transform How We Live,

Work, and Think. Eamon Dolan/Houghton Mifflin Harcourt 2013.

16. Editorial. Community cleverness required. Nature 2008; 455: doi:10.1038/455001a. 17. Souillard-Manda W. Davis, R, Rudin C, *et al*. Interpretable Machine Learning Models for the Digital Clock Drawing Test. International Conference of Machine Learning (ICML), New York, USA 2016.
18. Ioannidis JP. How to make more published research true. PLoS Med. 2014; 11: e1001747.
19. Eklund A, Nichols TE, Knutsson H. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad. Sci. USA 2016; 113: 7900-5.
20. Nuzzo R. Fooling ourselves. Nature 2015, 526: 182-5.